

VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations

Year: 2020 | Citations: 401 | Authors: H. Nguyen, K. Lam, Linh T. Le, Hieu Pham, Dat Q. Tran

Abstract

Most of the existing chest X-ray datasets include labels from a list of findings without specifying their locations on the radiographs. This limits the development of machine learning algorithms for the detection and localization of chest abnormalities. In this work, we describe a dataset of more than 100,000 chest X-ray scans that were retrospectively collected from two major hospitals in Vietnam. Out of this raw data, we release 18,000 images that were manually annotated by a total of 17 experienced radiologists with 22 local labels of rectangles surrounding abnormalities and 6 global labels of suspected diseases. The released dataset is divided into a training set of 15,000 and a test set of 3,000. Each scan in the training set was independently labeled by 3 radiologists, while each scan in the test set was labeled by the consensus of 5 radiologists. We designed and built a labeling platform for DICOM images to facilitate these annotation procedures. All images are made publicly available in DICOM format along with the labels of both the training set and the test set. Measurement(s) diseases and abnormal findings from chest X-ray scans Technology Type(s) AI is used to detect diseases and abnormal findings Sample Characteristic - Location Vietnam Measurement(s) diseases and abnormal findings from chest X-ray scans Technology Type(s) AI is used to detect diseases and abnormal findings Sample Characteristic - Location Vietnam